IGR204 - EuropeDisease

DELIN Jia - DILOU Sully - CALVET Rodolphe
TIZAF Zakaria - BINUANI Nicolas

June 2021

Table des matières

1 Introduction

The goal of this project is to give you some hands-on experience with creating a fully-functional interactive information visualization. This consists of several components : understanding the user of the visualization, what kinds of tasks he or she is trying to perform with the visualization, what is his or her background and expertise, and what kind of data we're dealing with.
For this project, you will choose a dataset to visualize. Explore the dataset, what attributes it contains, their types, how many cases are in the set, etc. Think about whether you will need to clean it up, complete the data with data from other sources, etc.
In this project, you will develop a high-quality, interactive visualization that will showcase your own interesting ideas. Be creative. Think about the data and what the user is trying to do with it. Think about appropriate representations. Think about appropriate interactions to transform the data or their representation. Most of all, be creative. And be creative.
This is not a programming project, but it is a project that will involve programming. Interaction plays an essential role in visualization, and it is difficult to understand the interaction in a visualization without a running system. You will be evaluated based on your creativity, how well you are able to describe the analytic tasks your system aims to support, and on the functionality of your visualization prototype. You will not be evaluated directly on the quality of the code you write.

2 Instructions

Your report should cover the following topics. I am looking for thoughtful reflection in your reports—do not merely answer the questions below, but also describe your rationale where appropriate. Think about why you made the choices you made and how you might articulate your design rationale. Were there other approaches that you ended up abandoning ?
The chosen data set What does it represent How many cases are there ? What are the dimensions, their properties, etc. (e.g. are they ordinal, nominal, etc.) Who are the target users ? What are they trying to understand from the data ? What kind of experience do they have with the data ? What kind of experience do they have with visualization tools ? What are some representative tasks ? What kinds of things would a user potentially want to understand from the data ? What questions helped guide your design ? What is your chosen design ? What representations did you choose to use ? What interaction does the design handle ? Can you relate these design decisions to your chosen users, data, and tasks ? What does the design do well ? What does it not do well ? A link to your git repository for the code. Be sure to include the revision number for your final version. Don't forget to give me access to the repository if it's private. I am eaganj on github, bitbucket, and gitlab. Reports will probably be somewhere around 10–20 pages, but that's not a target : use the space that you feel is justified to cover all of these topics. Feel free to write a longer report, especially if you include a lot of images. Be sure to use a legible, serif font at a reasonable size (10-12 point) with reasonable margins.

3 Dataset

Reminder on the topic, the datasets used, where the data come from, and their formats.
Dataset name : *Health_Europe_dataset_somediseases*
Source : `http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlth_cd_anr&lang=en`
Format : csv

Description of the users :

- Who ?
  OMS - Doctors - Pharmacy companies who want to follow the evolution of some disease depending on age, country or gender and apply or modify health plans.

- What are their backgrounds ?
  Health domain expertise and maybe some politicians.

- What are they trying to understand from the data ?
  The evolution of some diseases in each country and during a decade to validate or not some health policies taken (example with ALCOVE European program for Alzheimer disease 2011)

- Is your visualization aimed primarily at exploring or communicating the data ?
  The first goal is to communicate on the evolution of some diseases in Europe. After that, we can also use it to explore some new insights and change health European programs.

Now let's move to the dataset :

- What are the characteristics of the data, the attributes, the size of the dataset, etc. The dataset gives the total number of deaths by disease, gender, age, and year from 2001 to 2010 in Europe.
- Shape [190000 rows * 6 columns]

Here is an example of the tabular dataset. Columns Time (Year) and Value (Number) are integers, the other columns present textual data.

Year	GEO	SEX	AGE	Disease	Number
2001	Belgium	Female	From 25 to 29 years	Pneumonia	261

Also you can find the description of each column type :

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 179718 entries, 0 to 190079
Data columns (total 8 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   TIME               179718 non-null  int64
 1   GEO                179718 non-null  object
 2   UNIT               179718 non-null  object
 3   SEX                179718 non-null  object
 4   AGE                179718 non-null  object
 5   ICD10              179718 non-null  object
 6   Value              179718 non-null  int64
 7   Flag and Footnotes  4752 non-null    object
dtypes: int64(2), object(6)
memory usage: 12.3+ MB
```

FIGURE 1 – Description type of the dataset

4 Design

4.1 First approach

Our goal on this project is to have a tool for health experts, doctors, and politicians about evolution of some most known disease. In fact, we thought about CovidTracker application that is a little like a kind of dashboard to follow disease evolution. So because it's disease repartition between Europe countries, we first think about a chloropleth map that is obvious with this kind of problematic. This is a kind of representation that we want to use :

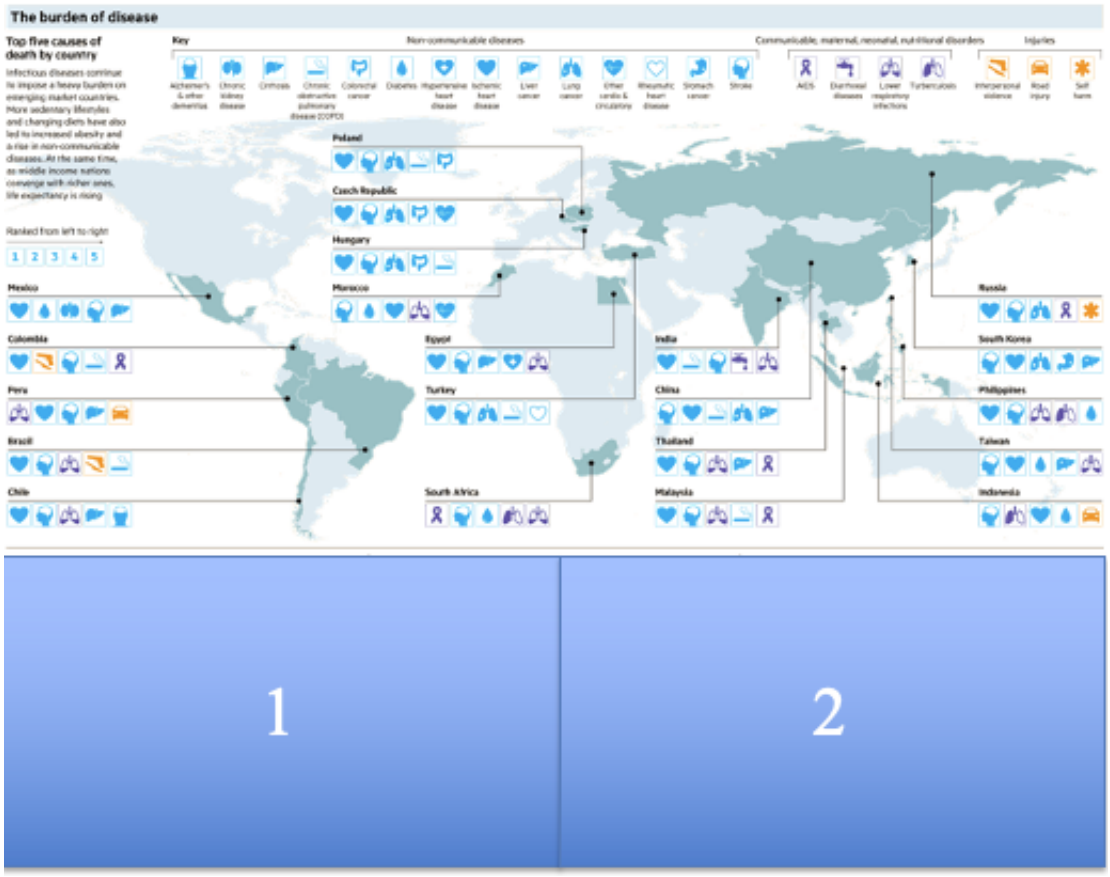

FIGURE 2 – Idea of dashboard for Europe disease

We have a map which we can select different countries and focus on some disease available in our database. Thus to explore more deeper we think about different plots.

4.2 Map design

In first case we think about a Europe map to represent disease.
For example, in our first approach, we suggest a map with disease most representaded in each country. See below for a first example of such representation.

FIGURE 3 – Idea of dashboard for Europe disease

A problematic in this first idea is taht some disease in our dataset are most represented in many countries, therefore maybe two or three diseases will be overexposed and it's not our goal. Remember we want to help politicians (OMS) and medical supplies get more insights, so we more want to focus on one disease and see the repartition between each country.
So now, we go deeper with adding other plots depending on disease selected.

FIGURE 4 – Idea of dashboard for Europe disease

Now let's start with our own drawn visualization. We begin with our map and add a ListBox.

- Solution1 Request1 : we have for a given disease, the evolution of count VS year on a EU map.
- Solution2 Request2 : we have for a given country, we show top 5 diseases with dynamics details.
- Solution3 Request3 : we have for a given age/slice of age, we show top 5 diseases

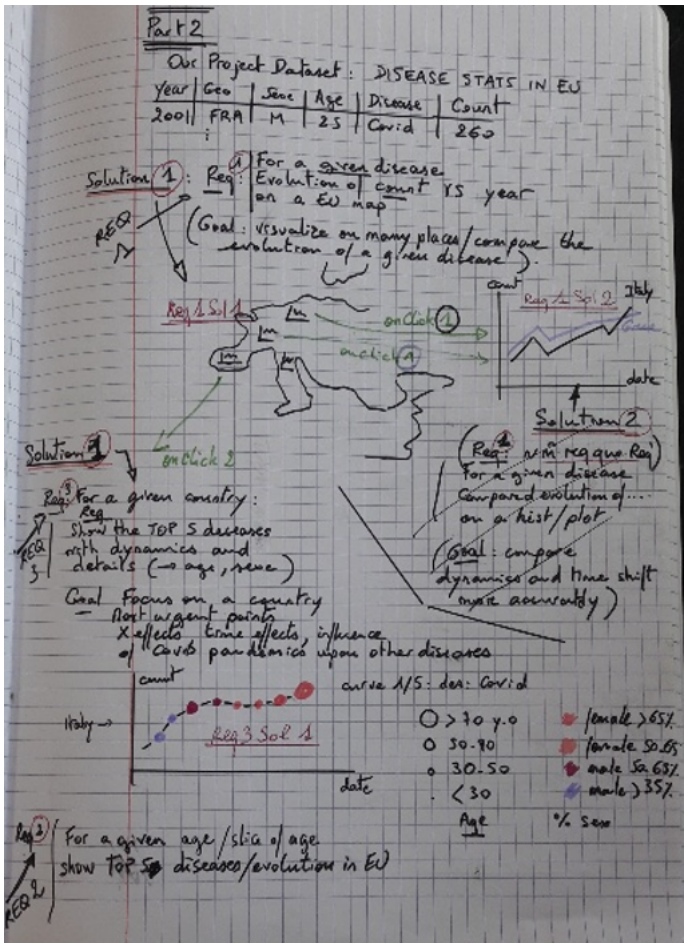

FIGURE 5 – Idea of dashboard for Europe disease

In a second approach, we think about add animated color legend like representing in following drawing sketchs :

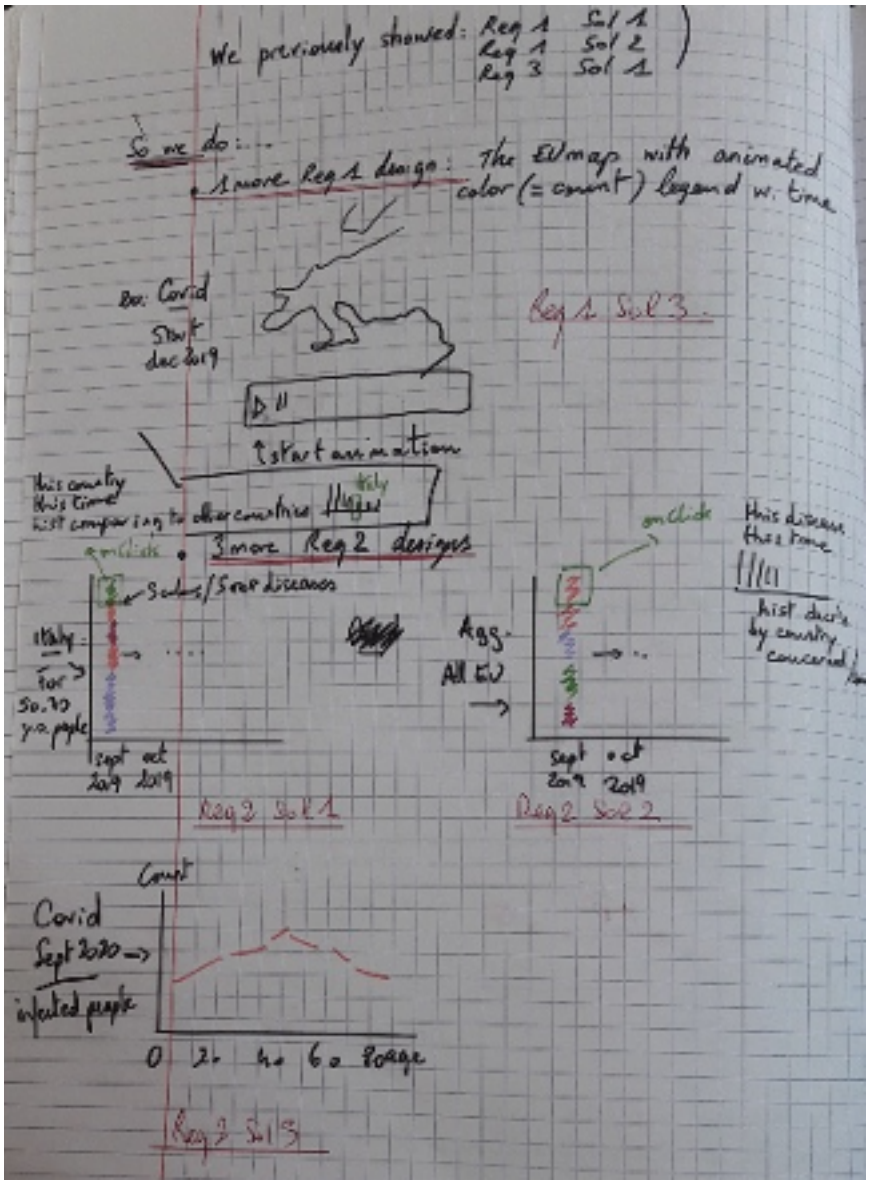

FIGURE 6 – Idea of dashboard for Europe disease

So to conclude on this first approaches, we decided to have a chloropleth map with a selected disease from a ListBox. We add 2 others plots like a dashboard visualization. After some Solution Request sketches we think for examples on the following plots :

— We can also have a barplot for one disease chosen and have a kind of

ranking with a barplot / or a rectangular diagram board to focus on which country is more affected (refer to part II.3)

- We can have gender like the example, and we can focus on age repartition depending on disease observed and for each country. (Refer to part II.3)
- We can select one country on a map and have a time serie plot of one disease also selected and for each year point size depending on age patient.
- We can select one country on a map and have evolution of count victims compare to mean of Europe indeed we have a kind of threshold
- Also, one point we think about is maybe to select one country on a map and have a kind of country disease pie chart such as follow.

So now let's move to a second section to focus on dashboard plots exploration and choose the 2 to add on our project.

4.3 Dashboard plots

Here is a kind of summary of which request and how we want it to interact with our visualizations.

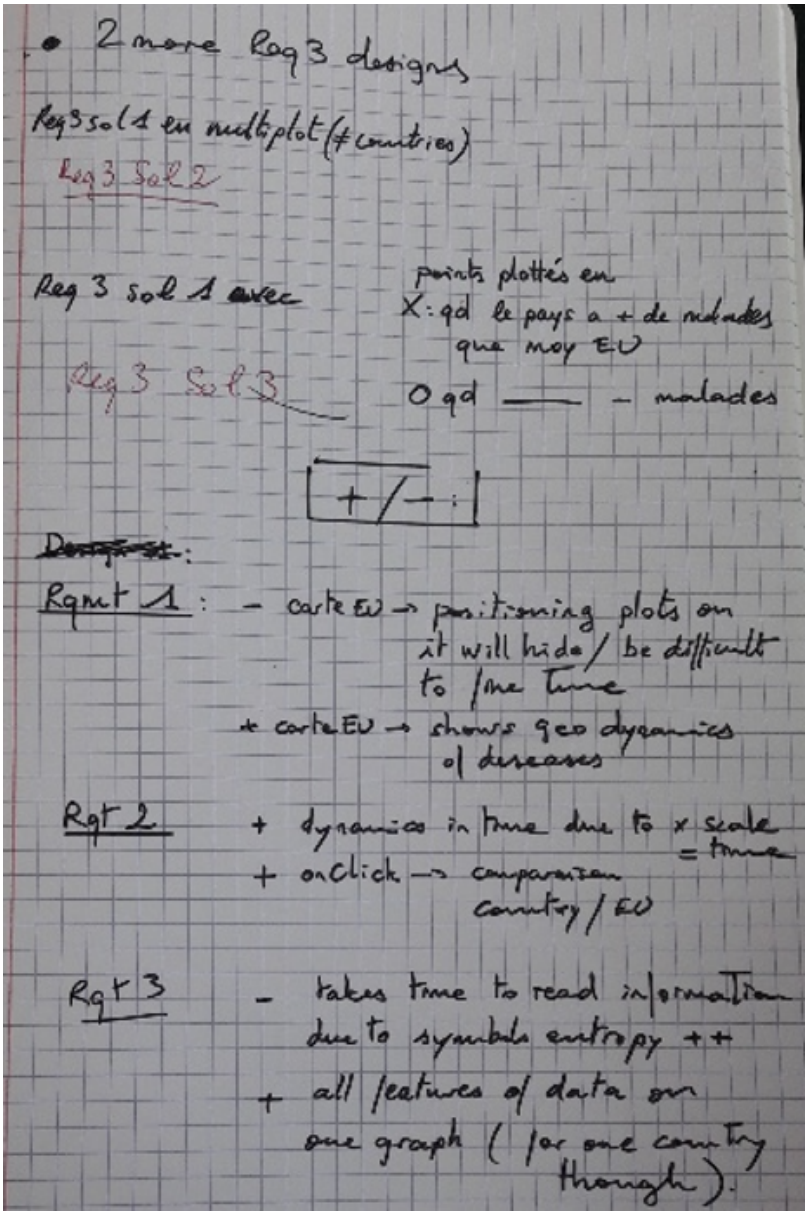

FIGURE 7 – Idea of dashboard for Europe disease

After our discussions in previous section, we first think about a treemap (Figure8) with all diseases but as we add a ListBox to choose the disease we want to focus, we turn treemap to a pyramid barplot to focus more about victim's characteristics (age, gender). So we turn this first idea from Figure8 to Figure9.

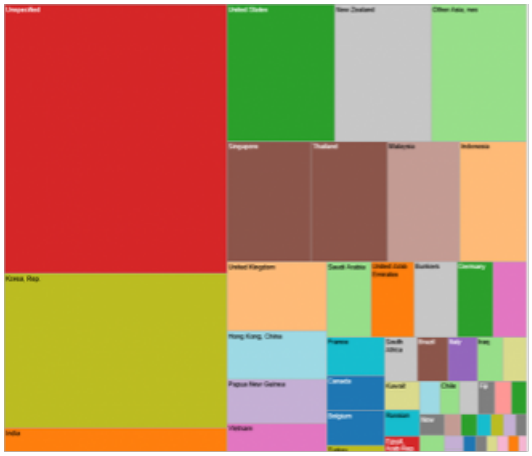

FIGURE 8 – Idea of dashboard for Europe disease

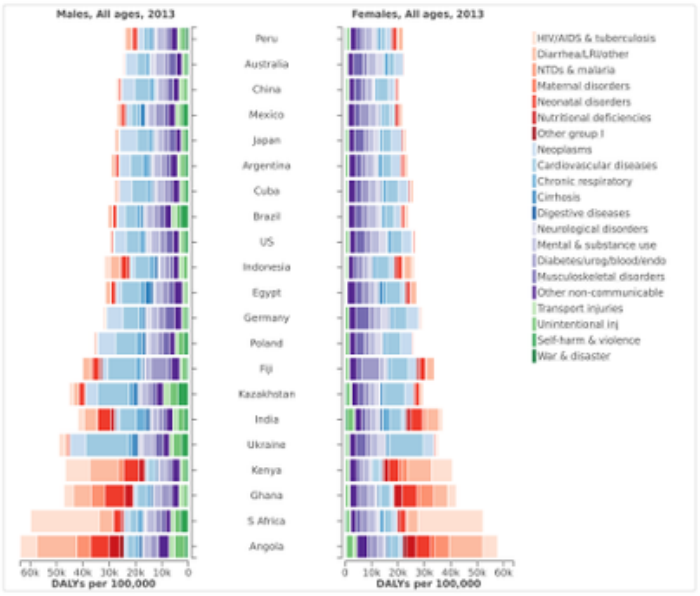

FIGURE 9 – Idea of dashboard for Europe disease

For the last graphic, we think that we missing a time representation to focus also on the evolution of one disease selected between all countries. So for that plot, we simply think about a time serie that represent each country disease evolution for 10 years range.

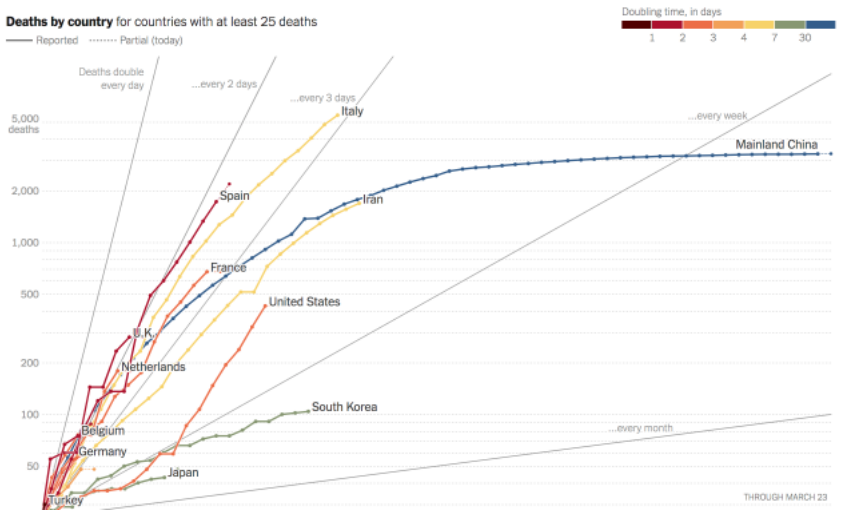

FIGURE 10 – Idea of dashboard for Europe disease

To conclude, after all sketches and our drawings, we focus on a ListBow to choose one disease we want to focus. After that, we have our choropleth map depending on the disease selected and 2 plots. One focus on victim age gender repartition and another one focus on the evolution of the disease in 10 years range.

Now let's explain our code and make a presentation.

5 Presentation

To create our visualization we need a tool enabling us to link multiple views of the data to show the several information in our data set. As we experienced during previous labs, Altair is a good tool for combining several visualizations with an interactive dimension. **Altair** is a declarative statistical visualization library for Python, based on Vega and Vega-Lite.
After assessing our respective abilities with Altair, we agreed on a simplified visualization that answers the questions presented above. Our visualization is composed of three different views. For each view, data values are bind to a specific cause of death that we can choose via a drop-down list.
Our first view is a map with a color scale encoding that shows the total number of deaths for the European countries on the cause we selected. It allows us to have a quick look at the difference of number of deaths between countries : the colder the color is, the more number of death is caused by this cause in the country. To create this view, we used the "geoshape marks" tool for European countries map.

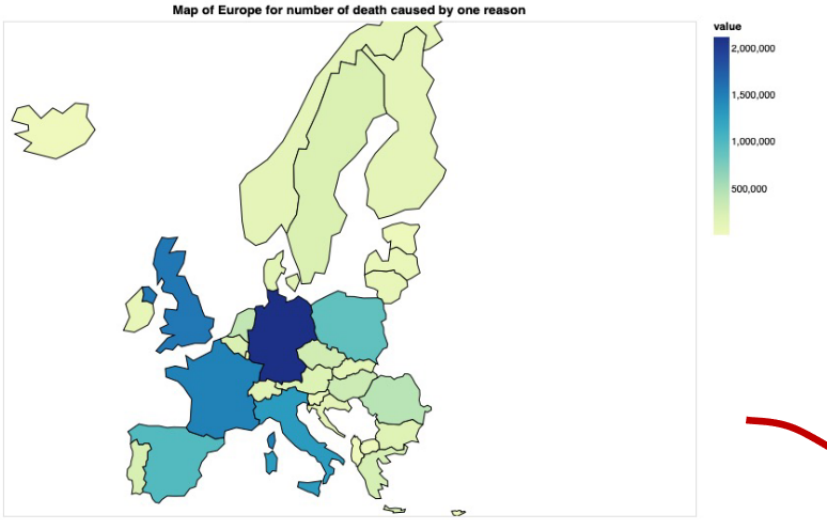

FIGURE 11 – First view

For this specific view, we needed to use an other data set and combine it with the first one. It contains country name and its corresponding **ISO 3166-1** numeric code for over 200 countries. It aims to visualise each country on the map for Altair.

Code	Country name
020	Andorra

The second view is a graph that shows the evolution of number of deaths due to one cause of death from 2001 to 2010. This graph helps us better understand the evolution of disease in each country and may improve the health plan for different countries. This view is also linked to the cause of death selected in the drop-down list.

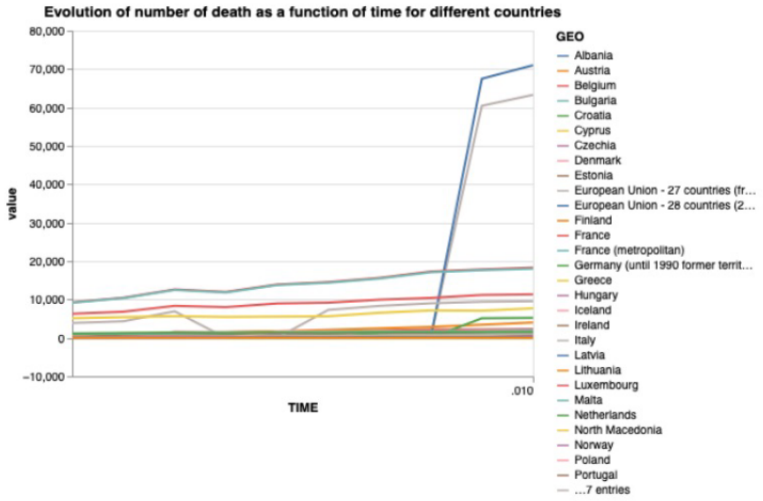

FIGURE 12 – Second view

The last view adds another aspect of our data. It allows us to compare between different genders. On the left we have the age distribution of number of death caused by one disease for male and on the right, the age distribution of number of death for female is shown. For example, we can see that the number of deaths caused by Alzheimer disease for female is higher the one for male, which can give the hint for hospitals and other institutions to pay more attention to female patient with Alzheimer.

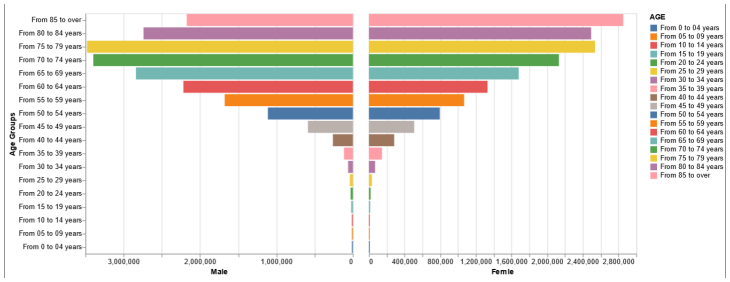

FIGURE 13 – Third view

6 Conclusion

This project was a good way for us to understand the process of creating a meaningful visualization from scratch. We also had a glimpse of the several difficulties we may encounter in such projects, from the task of defining the question we want to help to answer to the kind of tools we can use.
The fun part is to discover what data says to us once is encoded in a good visualization and we believe it represents an important motivation in the task of creating insightful visualization.